

Responsible NLP Checklist

Paper title: *Query-Aware Graph Attention for Precise Subgraph Retrieval in Knowledge-Augmented Reasoning*

Authors: *Yuanye Xu, Linyi Guo, Yue Zhang, Fu Ning*

How to read the checklist symbols:

- the authors responded 'yes'
- the authors responded 'no'
- the authors indicated that the question does not apply to their work
- the authors did not respond to the checkbox question

For background on the checklist and guidance provided to the authors, see the [Responsible NLP Checklist](#) page at ACL Rolling Review.

A. Questions mandatory for all submissions.

A1. Did you describe the limitations of your work?

This paper has a Limitations section.

A2. Did you discuss any potential risks of your work?

We discuss the potential risks in the "Ethics Considerations" section. While our method improves the precision of knowledge retrieval to mitigate hallucinations, the generated content still depends on the underlying Large Language Model and the quality of the external Knowledge Graph. We acknowledge that if the source KG contains biases or incorrect information, the system might reinforce these inaccuracies. We also emphasize that our model should be used as a reasoning assistant and not as a source of absolute factual truth in sensitive domains.

B. Did you use or create scientific artifacts? (e.g. code, datasets, models)

B4. Did you discuss the steps taken to check whether the data that was collected/used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect/anonymize it?

We used widely-adopted, publicly available benchmark datasets (WebQSP and CWQ) which are derived from Freebase. These datasets have been pre-processed by the original creators and the research community to ensure they contain factual knowledge graph information rather than personally identifying information or offensive content. Therefore, no additional anonymization steps were required for this study.

B6. Did you report relevant statistics like the number of examples, details of train/test/dev splits, etc. for the data that you used/created?

Section 5.1 (Experimental Setup). We report the statistics for all datasets used. Specifically, we use WebQuestionsSP (WebQSP) with 4,737 questions and Complex WebQuestions (CWQ) with 34,699 questions, both based on Freebase. We provide details on the reasoning hops (up to two-hop for WebQSP and higher compositionality for CWQ) and confirm that we follow the standard preprocessing and train/test splits as established in prior work (e.g., RoG).

C. Did you run computational experiments?

C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?

The Responsible NLP Checklist used at ACL Rolling Review is adopted from NAACL 2022, with the addition of ACL 2023 question on AI writing assistance and further refinements based on ARR practice. ACL 2026 used a subset of ARR checklist form.

Section 5.1 (Experimental Setup) and Appendix A (Additional Experiments). In the main text, we describe the choice of LLMs (Llama-3.1-8B-Instruct, GPT-4o-mini), the embedding models, and the retrieval parameters. In the Appendix, we provide comprehensive hyperparameter details.

- C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?

In Section 5.2, we provide a comprehensive comparison of retrieval and reasoning performance against various state-of-the-art baselines on WebQSP and CWQ datasets. In Section 6, we report statistics from ablation experiments to quantify the specific contribution of the Query-Relational Graph Attention Network (QR-GAT). We also include an Error Analysis (Appendix C) to provide transparency regarding the model's performance across different error categories.

D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?

- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?

This research does not involve human subjects, crowdworkers, or manual data annotation. All experiments were conducted using publicly available benchmark datasets and automated computational models. Therefore, no instructions were given to human participants.

- D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?

This study did not involve any human participants, crowdworkers, or recruitment processes. The evaluation was conducted entirely through automated experiments on publicly available benchmark datasets. As such, no recruitment or payment information is applicable.

- D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating (e.g., did your instructions explain how the data would be used)?

This research utilizes publicly available, third-party benchmark datasets (WebQuestionsSP and Complex WebQuestions) that were originally derived from Freebase. These resources are widely used in the NLP research community and have established terms of use for academic purposes. Since we did not collect new data from human subjects, the process of obtaining individual consent is not applicable to this study.

- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?

This study exclusively uses publicly available, anonymized benchmark datasets (WebQSP and CWQ) that are standard in the NLP community. These datasets were collected by previous researchers and do not contain personally identifiable information or involve active human participation in our study. Therefore, an ethics review board approval (IRB) was not required.

E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?

- E1. If you used AI assistants, did you include information about their use?

AI assistants (e.g., ChatGPT) were used for language polishing and clarity improvement during manuscript preparation and checklist completion. All content was reviewed and finalized by the authors.